Developing students' understanding of univariate statistical graphs: A local instruction theory for primary and secondary education

Lonneke Boels

HU University of Applied Sciences Utrecht

Although numerous studies have examined how students interpret individual graph types, research remains fragmented, lacking a coherent instructional theory on how to develop students' understanding of distribution. This paper addresses: How can students be supported in moving from case-based to distributional reasoning when interpreting statistical graphs? In response, a local instruction theory is proposed that articulates a progression of graph types—from case-value plots through dotplots to histograms and boxplots—aimed at bridging the conceptual gap between case-based and aggregated representations. This proposal is informed by eye-tracking and statistics education research showing that many students struggle to understand data aggregation in histograms and boxplots. The paper outlines implications for teaching materials aiming to foster distributional thinking across school levels.

Keywords: statistical literacy; histogram; boxplot; histodot; hatplot

Introduction

Statistical graphs are central to statistical literacy (Garfield & Ben-Zvi, 2008), yet students across educational levels experience persistent difficulties when interpreting distributions in graphs such as histograms and boxplots (Boels et al., 2019; Heursen et al., 2026). Existing research offers limited guidance on how students can be supported in moving from case-based to aggregated representations that require distributional reasoning. This paper addresses this gap by proposing a local instruction theory for supporting students' understanding of statistical graphs.

Recent eye-tracking studies reveal how students visually engage with different statistical graphs while solving tasks (Abt et al., 2024; Schreiter & Vogel, 2023). For histograms, students frequently confuse frequencies with individual values (Boels et al., 2025; Cooper, 2018) indicating that students do not understand the aggregated nature of data in histograms. For boxplots, some students may misinterpret a larger box as more data (Abt et al., 2024; Lem et al., 2013).

Taken together, these findings point to the need for a theoretically grounded perspective on the sequencing of graph representations that explicitly addresses the transition from case-based to distributional reasoning and that can inform research, instructional design and textbook development. The question for this theoretical paper is: *How can students be supported in moving from case-based to distributional reasoning when interpreting statistical graphs?*

Theoretical background

Reasoning about distribution involves viewing data as an aggregate characterised by centre and variability rather than as a set of individual cases (Garfield & Ben-Zvi, 2008). Many students struggle to make this shift and continue to rely on case-based interpretations when working with graphs of aggregated data (Boels et al., 2019; Heursen et al., 2026). For example, histograms are often interpreted as case-value plots, with bar heights treated as measured values rather than as frequencies of these within intervals (Boels, 2023; Boels et al., 2025). Frischemeier et al. (2023) identified several approaches for comparing distributions, including reasoning about centre, spread, shift, skewness, proportions below or above a given value (p-based), and quartiles (q-based). These forms of reasoning are conceptually demanding and depend on understanding data aggregation. Together, these findings suggest that distributional reasoning does not emerge automatically from introducing more abstract graphs. Instead, it requires instructional support that makes the data aggregation explicit.

In addition, eye-tracking research showed that students often focus on visually salient but conceptually misleading features of statistical graphs, such as bar heights or stacks, while overlooking intervals and aggregation (Boels, 2023; Boels et al., 2025; Boels & Van Dooren, 2023). It further indicates that comparing distributions across graphs is more demanding than interpreting a single graph, with difficulties persisting even when individual graphs appear to be understood (Boels et al., 2025; Schreiter & Vogel, 2023). Moreover, students may apply similar visual strategies across different graph types, even when these strategies are inappropriate, for example by interpreting case-value plots as histograms even when they attended to formal features such as axis labels (Boels et al., 2025). While the area is proportional to the data in histograms, in boxplots the area of the box is always representing the middle 50%, hence, a larger box is a larger interquartile range, which is often misinterpreted by students (Abt et al., 2024). This conceptual change of the meaning of the area in boxplots may be hard to grasp.

Although the body of research provides detailed insights into students' interpretations of individual graph types, it remains fragmented. Furthermore, textbooks often introduce histograms and boxplots as separate topics or as replacements for earlier representations, rather than as part of a coherent progression (Boels & Şeker, 2025). Moreover, dotplots and their variants as well as density histograms (Boels & Shvarts, 2023) are (almost) completely absent in Dutch Grades 7–12 textbooks (Boels & Şeker, 2025).

A local instruction theory for developing students' understanding of graphs

A local instruction theory is proposed for supporting students in progressing from case-based reasoning to distribution-based reasoning by foregrounding the role of intermediate representations and carefully designed progressions (Bakker, 2004). It articulates a progression of graph types—from case-value plots through dotplots to histograms and boxplots (figure 1). Gravemeijer (2004) described a local instruction theory as a theory that "refer[s] to the description of, and rationale for, the envisioned learning route as it relates to a set of instructional activities for a specific topic (e.g., addition and subtraction up to 20, area, fractions, etc.)" (p. 107).

Drawing on logical-historical analyses of statistical concepts, this approach emphasises that representations reify particular actions and ways of reasoning, which need to be unpacked and made accessible to learners (Boels et al., 2023). The local instruction theory treats histograms and boxplots not as replacements for earlier representations, but as aggregated forms that highlight the need for a conceptual transition from individual data points to aggregated distributions. Intermediate representations are introduced to support this transition explicitly. Dotplots are used as eye-tracking studies indicate that these invite students to attend to the horizontal axis and the underlying variable, while still keeping individual data points visible (Boels & Van Dooren, 2023). Therefore, dotplots and their variants can function as conceptual bridges between case-based and aggregated graphs, provided they are embedded in a coherent sequence rather than introduced in isolation (Boels, 2023, 2025a, 2025b). The local instruction theory starts with horizontal case-value plots because they are familiar to students and allow a transition towards dotplots through alignment of bar endpoints with the horizontal axis (Bakker, 2004).

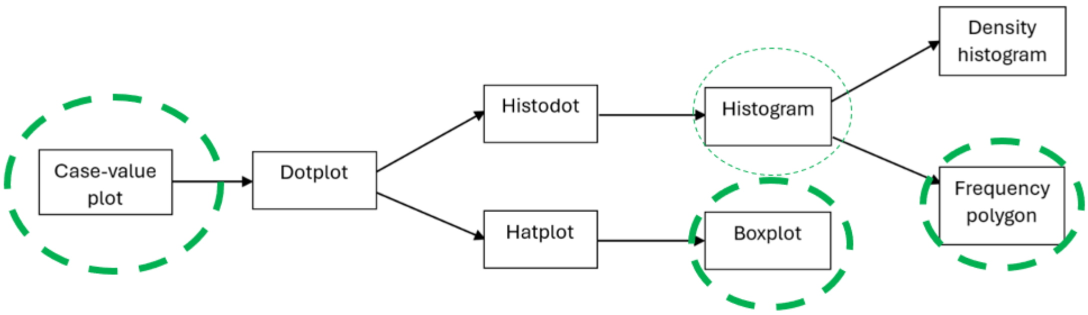

Figure 1. Overview of the proposed local instruction theory. The green dotted lines indicate graph types present in current Dutch textbooks. Dotplots are almost absent in most Dutch textbooks. All these graphs—except case-value plots—are univariate.

Two dotplot variants can be distinguished: one in which all dots are placed at their exact horizontal position and vertically displaced only to avoid overlap (sometimes referred to as 'messy' dotplots; Boels, 2023), and one in which dots are stacked within intervals (Schreiter & Vogel, 2024). Research indicates that stacked dotplots may induce misinterpretations similar to those observed for histograms both in primary and secondary education (Boels, et al., 2019; Schreiter & Vogel, 2024). Moreover, messy dotplots preserve the location of individual data points. When equal intervals are subsequently introduced in secondary education as an overlay to these dotplots, the resulting representations are histodots. At least two versions of histodots can be distinguished, each with its own merits (figure 2). The version used by Bakker (2004) foregrounds the role of equal intervals (figure 2, left), whereas the version used in VUstat.eu emphasises the individual measured values within each interval that together constitute the bar height (figure 2, right).

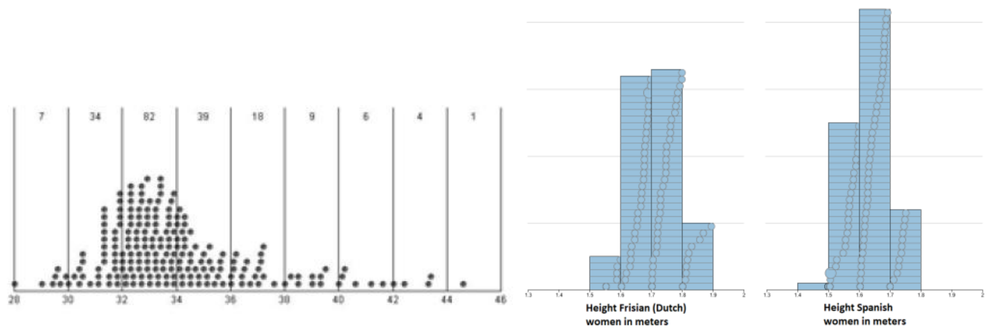

Figure 2. Histodot from Bakker (left , 2004, p.25, reprinted with permission) and histodot with histogram overlay created in https://www.vustat.eu/apps/stat/index.html?language=0 (right) by choosing a dotplot, the variable height, and ticking the histogram box.

In progressing from dotplots to boxplots, the division of data into equal portions becomes salient. When upper primary and lower secondary students describe characteristics of a distribution, they often refer to features such as a *bump* or *clump* (Bakker, 2001; Fielding-Wells, 2018). In software environments, this bump may be visualised as a hat, with the full interval indicated by a horizontal line (figure 3, left).

In subsequent representations, this hat becomes the middle 50% of the data. Adding the median completes the boxplot overlay (figure 3, right). This latter representation—a dotplot with boxplot overlay—is sometimes called a box-dotplot (Boels, 2025b). Dotplots, histodots and hatplots are included in the sequence because they allow distributional reasoning to develop gradually, each highlighting aspects of aggregation that remain implicit in histograms and boxplots themselves. Within this theory, histodots and hatplots function as endpoints for primary education, while secondary education is envisaged to begin with dotplots and their variants in lower grades and to extend to histograms and boxplots in upper grades.

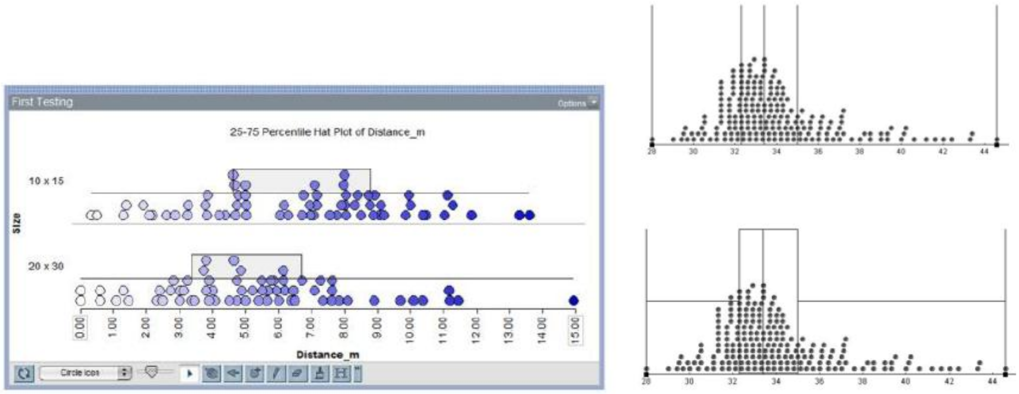

Figure 3. Hatplot made in TinkerPlots (left, Fielding-Wells, 2018, p. 1133, reprinted with permission) and adding quartiles (right, Bakker, 2004, p. 24).

Discussion and conclusion

This paper contributes to statistics education research by articulating a local instruction theory that shifts students' attention from individual cases to aggregated distributions through a coherent sequence of mostly univariate graphs (figure 1). A central theoretical insight is that aggregation fundamentally changes graph interpretation. Both histograms and boxplots represent aggregated data, which places substantial demands on learners. A key contribution of this paper is the argument that aggregation itself, rather than the specific graph type, constitutes a major conceptual hurdle in interpreting histograms and boxplots.

The proposed local instruction theory makes this transition to aggregated graphs explicit by positioning dotplots, histodots, and hatplots as intermediate representations. Non-stacked 'messy' dotplots keep the individual values visible while already supporting attention to the horizontal axis and the distribution as a whole. Histodots make the transition to histograms explicit by introducing equal intervals in dotplots and prepare for proportion-based reasoning (cf. Frischemeier et al., 2023). Hatplots foreground the idea of searching for the middle data part of a distribution in dotplots, thereby preparing learners for quartile-based reasoning (cf. Frischemeier et al., 2023).

These insights have direct implications for lesson materials and textbook design. Analyses of current textbooks suggest that the main issue is the absence of explicit representational progressions (Boels & Şeker, 2025). Aggregated graphs are often introduced as self-contained topics or as replacements for earlier representations, leaving learners to infer for themselves what has changed conceptually. The local instruction theory proposed here provides a research-informed basis for designing such progressions in research and lesson materials, making visible which ideas are being introduced, abstracted, or suppressed at each step.

Finally, the local instruction theory is intended as an analytic and design-oriented framework rather than a tested instructional sequence. As Gravemeijer (2004) stated: “The idea is that the teachers use their insight in the local instruction theory to choose instructional activities and to design hypothetical learning trajectories for their own students (p. 107).” Further research is needed to examine how these progressions function in classroom practice, how teachers take them up, and how they interact with curricular constraints. By articulating the conceptual role of aggregation, this paper contributes to a more precise theoretical understanding of students’ difficulties with statistical graphs and how instruction might better support them.

References

Abt, M., Leuders, T., Loibl, K., Strohmaier, A. R., Van Dooren, W., & Reinhold, F. (2024). How can eye-tracking data be used to understand cognitive processes when comparing data sets with box plots? *Frontiers in Education, 9.* https://doi.org/10.3389/feduc.2024.1425663

Bakker, A. (2001). Symbolizing data into a ‘bump’. In M. van den Heuvel-Panhuizen (Ed.). *Proceedings of the 25th Conference of the International Group of PME, Vol. 1* (pp. 81–88). Utrecht, The Netherlands: Freudenthal Institute. https://www.igpme.org/wp-content/uploads/2019/05/PME25-2001-Utrecht.zip

Bakker, A. (2004). Design research in statistics education: On symbolizing and computer tools [Doctoral dissertation, Utrecht University]. https://dspace.library.uu.nl/handle/1874/893

Boels, L. (2023). Histograms. An educational eye [Dissertation. Utrecht University]. https://dspace.library.uu.nl/handle/1874/430641

Boels, L. (2025a). Oogbewegingsonderzoek naar diagrammen. Pleidooi voor gebruik van stippendiagrammen op de basisschool [Eye-tracking research on diagrams. Advocating the use of dotplots in primary education]. *Volgens Bartjens – ontwikkeling en onderzoek, 44*(4), 41–50. https://www.volgens-bartjens.nl/media/10/vboo_44_04_01_maart25_boels.pdf

Boels, L. (2025b). Developing students’ understanding of statistical graphs. Local instruction theory for primary and secondary education. Poster presented at the *14th Congress of the European Society for Research in Mathematics Education (CERME14),* Free University of Bozen-Bolzano, Italy and ERME. https://doi.org/10.13140/RG.2.2.36750.68164

Boels, L., Alberto, R. & Shvarts, A. (2023). Actions behind mathematical concepts: A logical-historical analysis. In P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the 13th Congress of the European Society for Research in Mathematics Education (CERME13)* (pp. 5083–5090). Alfréd Rényi Institute of Mathematics and ERME. https://hal.science/CERME13/hal-04420753v1

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2019). Conceptual difficulties when interpreting histograms: A review. *Educational Research Review, 28,* Article 100291. https://doi.org/10.1016/j.edurev.2019.100291

Boels, L., Bakker, A., Van Dooren, W., & Drijvers, P. (2025). Secondary school students’ strategies when interpreting histograms and case-value plots: An eye-tracking study. *Educational Studies in Mathematics, 118,* 479–503. https://doi.org/10.1007/s10649-024-10351-3

Boels, L., & Şeker, V. (2025). The role of dotplots in statistics education practice. *IASE satellite conference. Statistics and Data Science Education in STEAM,* Münster, Germany. https://uni-muenster.sciebo.de/s/PnA9Krr44ZRPcGf

Boels, L., & Shvarts, A. (2023). Introducing density histograms to Grades 10 and 12 students. Design and tryout of an intervention inspired by embodied instrumentation. In G. F. Burrill, L. de Oliveria Souza & E. Reston (Eds.) *Research on reasoning with data and statistical thinking: International perspectives. Advances in Mathematics Education* (pp. 143–167). Springer. https://doi.org/10.1007/978-3-031-29459-4_14

Boels, L., & Van Dooren, W. (2023). Secondary school students interpreting and comparing dotplots: An eye-tracking study. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel & M. Tabach (Eds.). *Proceedings of the 46th Conference of the International Group for the Psychology of Mathematics Education, 2*, (pp. 123–130). PME. Haifa, Israel. https://www.researchgate.net/publication/372448356_Secondary_school_students_interpreting_and_comparing_dotplots_An_eye-tracking_study

Cooper, L. L. (2018). Assessing students' understanding of variability in graphical representations that share the common attribute of bars. *Journal of Statistics Education, 26*(2), 110–124. https://doi.org/10.1080/10691898.2018.1473060

Fielding-Wells, J. (2018). Dot plots and hat plots: Supporting young students emerging understandings of distribution, center and variability through modeling. *ZDM Mathematics Education 50*, 1125–1138. https://doi.org/10.1007/s11858-018-0961-1

Frischemeier, D. (2019). Primary school students' reasoning when comparing groups using modal clumps, medians, and hatplots. Mathematics Education Research Journal, 31, 485–505.

Frischemeier, D., Podworny, S., & Biehler, R. (2023). Data visualization packages for non-inferential civic statistics in high school classrooms. In J. Ridgway (Ed.), *Statistics for empowerment and social engagement: Teaching Civic Statistics to develop informed citizens.* Springer. https://doi.org/10.1007/978-3-031-20748-8_9

Garfield, J. B., & Ben-Zvi, D. (2008). Developing students' statistical reasoning: *Connecting research and teaching practice.* Springer. https://doi.org/10.1007/978- 1-4020-8383-9

Heursen, A., Schreiter, S., Şeker, V., Alexander, A., Boels, L., Reinhold, F., Abt, M., Van Dooren, W., & Vogel, M. (2026). *Review of systematic errors when interpreting statistical graphs.* [Manuscript submitted for review]. University of Applied Sciences, Heidelberg.

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013). On the misinterpretation of histograms and box plots. *Educational Psychology, 33*(2), 155–174. https://doi.org/10.1080/01443410.2012.674006

Schreiter, S., & Vogel, M. (2023). Eye-tracking measures as indicators for a local vs. global view of data. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.1058150